

## **Supplementary Information**

### **5-hydroxymethylcytosine-mediated epigenetic dynamics during postnatal neurodevelopment and aging**

By

Keith E. Szulwach<sup>1,8</sup>, Xuekun Li<sup>1,8</sup>, Yujing Li<sup>1</sup>, Chun-Xiao Song<sup>2</sup>, Hao Wu<sup>3</sup>, Qing Dai<sup>2</sup>, Hasan Irier<sup>1</sup>, Anup K. Upadhyay<sup>4</sup>, Marla Gearing<sup>5</sup>, Allan I. Levey<sup>5</sup>, Aparna Vasanthakumar<sup>6</sup>, Lucy A. Godley<sup>6</sup>, Qiang Chang<sup>7</sup>, Xiaodong Cheng<sup>4</sup>, Chuan He<sup>2</sup>, Peng Jin<sup>1\*</sup>

**Supplementary Table 1. Summary of sequencing data.** This table provides a summary of the number of biological replicates used per sample as well as the total number of aligned and monoclonal reads per sample. Sequence data is accessible through NCBI GEO Accession 32050.

	Biological Replicates	Total reads	Aligned	Monoclonal
<b>Mouse Brain</b>				
P7.Cb	2	45536567	32282873	28005945
6wk.Cb	2	48082397	34825975	28904372
1yr.Cb	2	48085069	34920333	24362956
P7.Hipp	2	57671616	41958357	33517754
6wk.Hipp	3	112017004	76783055	48142452
1yr.Hipp	2	53684252	37684943	29079892
Input	2	45253385	26831158	25980384
Mecp2-/ <i>y</i> <sup>tm1.1Bird</sup> Cb	2	50938891	39511650	30643116
<b>Human Cerebellum</b>				
Male Cb	1	46913108	28469364	13663115
Male Input	1	24039074	13605574	12699468
Female Cb	1	56028073	33186906	9342231
Female Input	1	27426391	15504818	13803442
<b>Rat Brain</b>				
6-8wk brainstem	1	47785386	29864254	10619796

**Supplementary Table 2. qPCR primer sets.** This table shows the list of qPCR primer sets used for validation of 5-hmC enriched regions and qPCR of MeDIP DNA.

<b><u>Tissue Specific qPCR</u></b>	
<u>En2-FW</u>	CCACTCTGGCAGCACGAA
<u>En2-RV</u>	TTTAGCCCCGGTGAGCTTGTC
<u>Nr2e1-FW</u>	CCCCACACAAGTGGCTTT
<u>Nr2e1-RV</u>	TCCCATCCCCCAGCTTTC
<b><u>MeDIP qPCR</u></b>	
<u>Exph5-FW</u>	TCAACCTGGGTGCACTTAACC
<u>Exph5-RV</u>	AACATAGTGCTGTGTACAGTGGACAA
<u>Pip5k1b-FW</u>	TGCTCACCCCGCTTGTG
<u>Pip5k1b-RV</u>	AGGCTGATAAGATAAACAGGTCATCA
<u>Grin2c-FW</u>	TGCCTTCCCTGGTCACATC
<u>Grin2c-RV</u>	GGGTGTAGGGCTGAGTTTGAAT
<u>Mapk12-FW</u>	CCAAGGCTCTGGGATGATAGAG
<u>Mapk12-RV</u>	AGACATACAGGATCCTAGGCATTCA
<u>BC145649-FW</u>	CTCAGGGCCCAGAGTTCGT
<u>BC145649-RV</u>	GAATTGGGTTCTCTAGGAGGTAAGT
<u>Dgkg-FW</u>	CGTTCCCTCTCATTCGTTACG
<u>Dgkg-RV</u>	CAAAGTGAAGAAGTTAGGAGATCAAGGA
<u>Il16-FW</u>	ATGCTAGGAAGCTGTGATATGTGAA
<u>Il16-RV</u>	TGTCTTTGGCCATTAGATCCA
<u>Gabrd-FW</u>	CCAGAAGGCATGGGAAGGTA
<u>Gabrd-RV</u>	TTATCCGCCTACAGCCTGATG
<u>Tesc-FW</u>	TGTGAGTTGAACCAACCTTTGC
<u>Tesc-RV</u>	TGTCATTGCTACATCGTAACTGTAATTT
<u>St8sia5-FW</u>	TGGCATCGAGGATTTTGA
<u>St8sia5-RV</u>	CACCGTACATGACAATCAGAAAGTT
<u>Psd3-FW</u>	CGTCTTTCCACCTCAGTTGATTC
<u>Psd3-RV</u>	CCACACATTTGCTGAGAGAAGCT
<u>Syt1-FW</u>	GAAATAGATAGTAGCAGGGTGGAGAAC
<u>Syt1-RV</u>	GGATGCCCTTTGGACATAGC
<b><u>Repeat 5-hmC qPCR validation</u></b>	
<u>IAPLTR1a_Mm-FW</u>	GTGCAGCATGGTGTAGTGTAGCT
<u>IAPLTR1a_Mm-RV</u>	CCTACACAAACAAGCAAACAAATGA
<u>B1_Mus2-FW</u>	GCCCCGTTGGTAGAGACATG

B1_Mus2-RV	TCCCCTGGCTCTCAATGG
B1_Mus1-FW	CCCACCACAGCCTGTTATACTTC
B1_Mus1-RV	GGGTGCCTTTTGCACCTTG
RSINE1-FW	GAATTGAACCCAGTCAGTCTTGGT
RSINE1-RV	CAAAATCCAATTTACAGAAGCTGCTA
B4A-FW	TGGGCGTCCCGTCTCA
B4A-RV	TGGAGGAGCTTTGGGTCTGA
PB1D9-FW	AGCGGCAGTGTCGAGGAA
PB1D9-RV	AACAAGGGAGGACTTGCATCTTAA
RLTR4_MM-int-FW	GGCTAAGGGACAATATTCCATCA
RLTR4_MM-int-RV	CCCCTTGACGGCAGTAGGT
MER3-FW	GAGGCAGGACCAAGCCTTTT
MER3-RV	CAAGCTATGCAGAAAGGACATCA
GSAT_MM-FW	GACGACTTGAAAAATGACGAAATC
GSAT_MM-RV	CATATTCCAGGTCCTTCAGTGTGC

## Supplementary Figure Legends

**Supplementary Figure 1. The expression and quantification of 5-hmC during neurodevelopment and aging.** **a.** Photomontage images showing the expression of 5-hmC in cerebellum across three different ages: P7 (left), 6-week (middle), and 1-year (right). **b.** Photomontage images showing the expression of 5-hmC in hippocampus across three different ages: P7 (left), 6-week (middle), and 1-year (right). **c.** Neural stem/progenitor cells (Nestin<sup>+</sup>) display either weak 5-hmC or a complete lack of 5-hmC. Scale bar: 50  $\mu$ m. **d.** Dot blot quantification of 5-hmC in cerebellum at P7, 6-week, and 1-year. Images are derived from the same blot and are representative of multiple independent blots. **e.** Dot blot quantification of 5-hmC in hippocampus at P7, 6-week, and 1-year. Images are derived from the same blot and are representative of multiple independent blots.

**Supplementary Figure 2. The expression of 5-hmC in the rostral migratory stream and olfactory bulb during neurodevelopment.** At P7, the anterior region of the subventricular zone (SVZa) is active neurogenic, and SVZa-derived neural stem cells migrate towards olfactory bulb (OB) along the rostral migratory stream (RMS). Neural stem cells become more fate-committed along migration. In the RMS-OB, the majority of cells are TuJ1 positive, but still NeuN negative. **a.** Low 5-hmC intensity is observed in SVZa at P7. Upper: DAPI, middle: 5-hmc; lower: NeuN. **b.** 5-hmC is enriched in the granular cell layer, where neuronal cells become postmitotic and become mature (NeuN<sup>+</sup>). Left: 5-hmC, right: NeuN.

**Supplementary Figure 3. Heatmap representation of input normalized 5-hmC signals in genome-wide binned (10kb) data in cerebellum and hippocampus at each age.** Input normalized 5-hmC values per 10kb were determined at each age for each brain regions and clustered using an uncentered correlation with pairwise complete linkage (Cluster3.0). 5-hmC values are equally scaled for all chromosomes, and chromosome number is indicated above the appropriate heatmap.

**Supplementary Figure 4. Metagene analysis of 5-hmC read distributions.** **a.** Read densities around all RefSeq transcripts summed in 10-bp bins centered on either TSS or TTS at each age in cerebellum. **b.** Read densities around all RefSeq transcripts summed in 10-bp bins centered on either TSS or TTS at each age in hippocampus.

**Supplementary Figure 5. Gene expression dependent 5-hmC densities in P7 and 6wk cerebellum.** **a.** 5-hmC densities  $\pm 5\text{kb}$  of TSS and TTS shown separately for P7 and 6wk Cb as a function of expression level. mRNA expression levels were determined by Affymetrix microarray and ranked in descending order by RMA normalized signal intensity. 5-hmC read densities are determined as the number of reads in 100bp bins normalized to the total number of reads in millions and divided by bin size. Scale is indicated in bottom right-hand corner **b.** The change in 5-hmC ( $\Delta 5\text{hmC}$ )  $\pm 5\text{kb}$  of TSS and TTS as a function of the change in gene expression ( $\Delta \text{Exp}$ ). All genes detected by microarray were ranked based on their change in RMA normalized signal intensity, determined as 6wk - P7, and ranked in descending order.  $\Delta 5\text{hmC}$  is determined as the difference in normalized 5-hmC read densities (6wk - P7) described above. Indicated to

the left are the regions corresponding to activated, unchanged, or repressed genes in 6wk Cb relative to P7 Cb. **c.** Quantification of values presented in **b.**, summarizing average  $\Delta 5\text{hmC}$  values  $\pm 5\text{kb}$  of TSS and TTS for activated, unchanged, or repressed genes, as determined by ANOVA, Benjamini-Hochberg adjusted p-value threshold of  $<0.05$  ( $n=3$ ). **d.** The cumulative fraction of genes at or below a given  $\Delta 5\text{hmC}$  value, considering only intragenic  $\Delta 5\text{hmC}$  at activated, unchanged, or repressed genes.

**Supplementary Figure 6. MeDIP analysis at developmentally activated genes with increased intragenic 5-hmC.** **a.** qPCR quantification of the fold change in 5-mC at ten 5-hmC regulated regions within genes that exhibit increased gene expression and increased 5-hmC in 6wk Cb relative to P7 Cb. Below the average O/E CpG dinucleotide frequency for the ten regions assayed is indicated. Shown to the right are four control regions. Two regions are within genes that do not exhibit altered expression and contain similar levels of 5-hmC enrichment (determined by 5-hmC peak p-value). The two regions to the far right are regions that do not exhibit 5-hmC enrichment at either age. Data were determined in two independent samples per age and are calibrated to a single P7 Cb sample, with fold-change calculated as  $2^{-\Delta\text{Ct}}$ . **b.** Summary of genes in **a.** showing the average fold-change in 5-mC determined by MeDIP for genes exhibiting increased gene expression and increased 5-hmC in 6wk Cb relative to P7 Cb.

**Supplementary Figure 7. Depletion of 5-hmC on chrX in rat brain and human male and female cerebellum.** **a.** Summary of rat chromosome-wide read densities showing depletion of 5-hmC on chrX, relative to autosomes. Normalized chromosome-wide read

densities were compared to those expected by chance assuming reads were randomly distributed amongst chromosomes. Dotted line represents a value of 1, or no enrichment. **b.** Summary of human chromosome-wide read densities showing depletion of 5-hmC on chrX in male and female human cerebellum. Shown is average 5-hmC enrichment on all autosomes compared to the enrichment observed on chrX. Dotted line represents a value of 1, or no enrichment. **c.** Genomic view of 5-hmC in male and female mouse cerebellum at two genes escaping X-inactivation, *Xist* and *Utx*. Shown is a heatmap representation of normalized read coverage in 25bp windows.

**Supplementary Figure 8. Sequence motifs, GC content, and CpG frequency associated with 5-hmC-regulated regions in mouse cerebellum and hippocampus.** **a.** For the union of all detected mouse DhMRs, the fractional GC content and the observed-to-expected (O/E) CpG frequency was determined and then compared to that of either CpG islands or randomly chosen genomic locations. **b.** For all detected regions enriched for 5-hmC in human cerebellum, the fractional GC content and the observed-to-expected (O/E) CpG frequency was determined and then compared to that of either CpG islands or randomly chosen genomic locations. **c.** Sequence motif associated with the top 50 (based on MACS p-value) cerebellum-specific DhMRs identified by MEME (E-value =  $5.8e-7$ ). **d.** Sequence motif associated with the top 50 (based on MACS p-value) hippocampus-specific DhMRs identified by MEME (E-value =  $6.3e-16$ ).

**Supplementary Figure 9. Examples of Dynamic and stable DhMRs in mouse cerebellum and Hippocampus.** **a.** Genomic view of a P7 Dynamic DhMR in cerebellum



demonstrating the presence of 5-hmC in P7 and subsequent loss at 6wk which is maintained through 1yr. **b.** Genomic view of a P7 Dynamic DhMR in hippocampus demonstrating the presence of 5-hmC in P7 and subsequent loss at 6wk which is maintained through 1yr. **c.** Genomic view of a Adult stable DhMR in cerebellum demonstrating the absence of 5-hmC in P7, but acquisition at 6wk and maintenance through 1yr. **d.** Genomic view of a Adult stable DhMR in hippocampus demonstrating the absence of 5-hmC in P7, but acquisition at 6wk and maintenance through 1yr. **e.** Genomic view of an Adult Dynamic DhMR in cerebellum demonstrating acquisition of 5-hmC from P7 to 6wk, but subsequent loss at 1yr. **f.** Genomic view of an Adult Dynamic DhMR in hippocampus demonstrating acquisition of 5-hmC from P7 to 6wk, but subsequent loss at 1yr. **g.** Genomic view of an DhMR specific to 1yr cerebellum demonstrating acquisition of 5-hmC at 1yr relative to both P7 and 6wk. **h.** Genomic view of an DhMR specific to 1yr hippocampus demonstrating acquisition of 5-hmC at 1yr relative to both P7 and 6wk. For all views, identified DhMRs are indicated by solid colored bars below the Input track.

**Supplementary Figure 10. qPCR validation of 5-hmC enrichment and dynamic regulation**

**at major repeat classes. a.** qPCR quantification showing enrichment of 5-hmC at five subtypes of SINEs, an LTR, and depletion of 5-hmC in major Satellite DNA in P7, 6wk, and 1yr cerebellum. **b.** qPCR quantification showing enrichment of 5-hmC at two subtypes of LTRs, SINEs, and depletion of 5-hmC in major Satellite DNA in P7, 6wk, and 1yr Hippocampus. Values are determined as the average of two independent samples by  $2^{-\Delta Ct}$ , with  $\Delta Ct$  calculated as the change in Ct value relative to input DNA.

**Supplementary Figure 11. *In vitro* 5-mC hydroxylation by Tet1-CD is blocked by MeCP2-**

**MBD.** To test the effect of MeCP2-MBD (aa97-167) on Tet1-CD (aa1367-2093) mediated hydroxylation of 5-mC, full 5-mC DNA (corresponding to 68 pmoles 5-mC) was held constant and MeCP2-MBD was varied to produce molecular ratios (MeCP2-MBD to 5mC) of 1:4, 1:2, 1:1, 2.5:1, 5:1, and no MeCP2-MBD), while also keeping the total amount of protein in the reaction constant by supplementing decreasing amounts of MeCP2-MBD with BSA. 5-hmC was quantified by 5-hmC specific immunoblot as described above, relative to the BSA only/no MeCP2-MBD control (mean  $\pm$  SEM, n=3). Shown below the x-axis is an immunoblot representative of replicates.

**Supplementary Figure 12. Summary of Tet family mRNA expression levels in P7 and 6wk**

**cerebellum.** Shown is the RMA normalized average signal intensity at P7 and 6wk Cb determined by Affymetrix microarray (n=3, mean  $\pm$  SEM).

## **Supplementary Datasets**

**Supplementary Dataset 1. P7 and 6 week Cerebellum gene expression data.** This dataset contains Affymetrix GeneChip Mouse Exon 1.0 ST array data from P7 and 6wk Cerebellum.

**Supplementary Dataset 2. P7 Cerebellum DhMRs.** This dataset contains DhMRs identified as enriched for 5-hmC by pairwise comparisons to 6wk and 1yr cerebellum as well as P7, 6wk, and 1yr hippocampus.

**Supplementary Dataset 3. 6 week Cerebellum DhMRs.** This dataset contains DhMRs identified as enriched for 5-hmC by pairwise comparisons to P7 and 1yr cerebellum as well as P7, 6wk, and 1yr hippocampus.

**Supplementary Dataset 4. 1 year Cerebellum DhMRs.** This dataset contains DhMRs identified as enriched for 5-hmC by pairwise comparisons to P7 and 6wk cerebellum as well as P7, 6wk, and 1yr hippocampus.

**Supplementary Dataset 5. P7 Hippocampus Cerebellum DhMRs.** This dataset contains DhMRs identified as enriched for 5-hmC by pairwise comparisons to 6wk and 1yr hippocampus as well as P7, 6wk, and 1yr cerebellum.

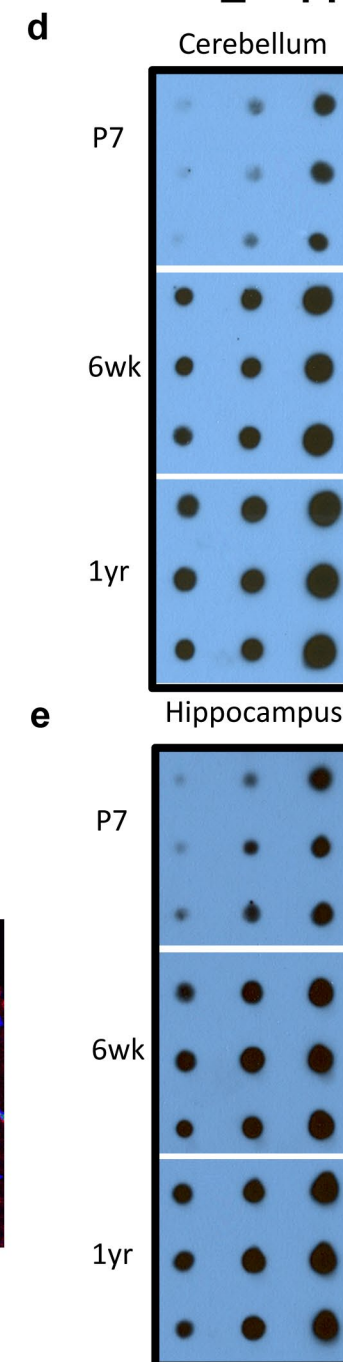
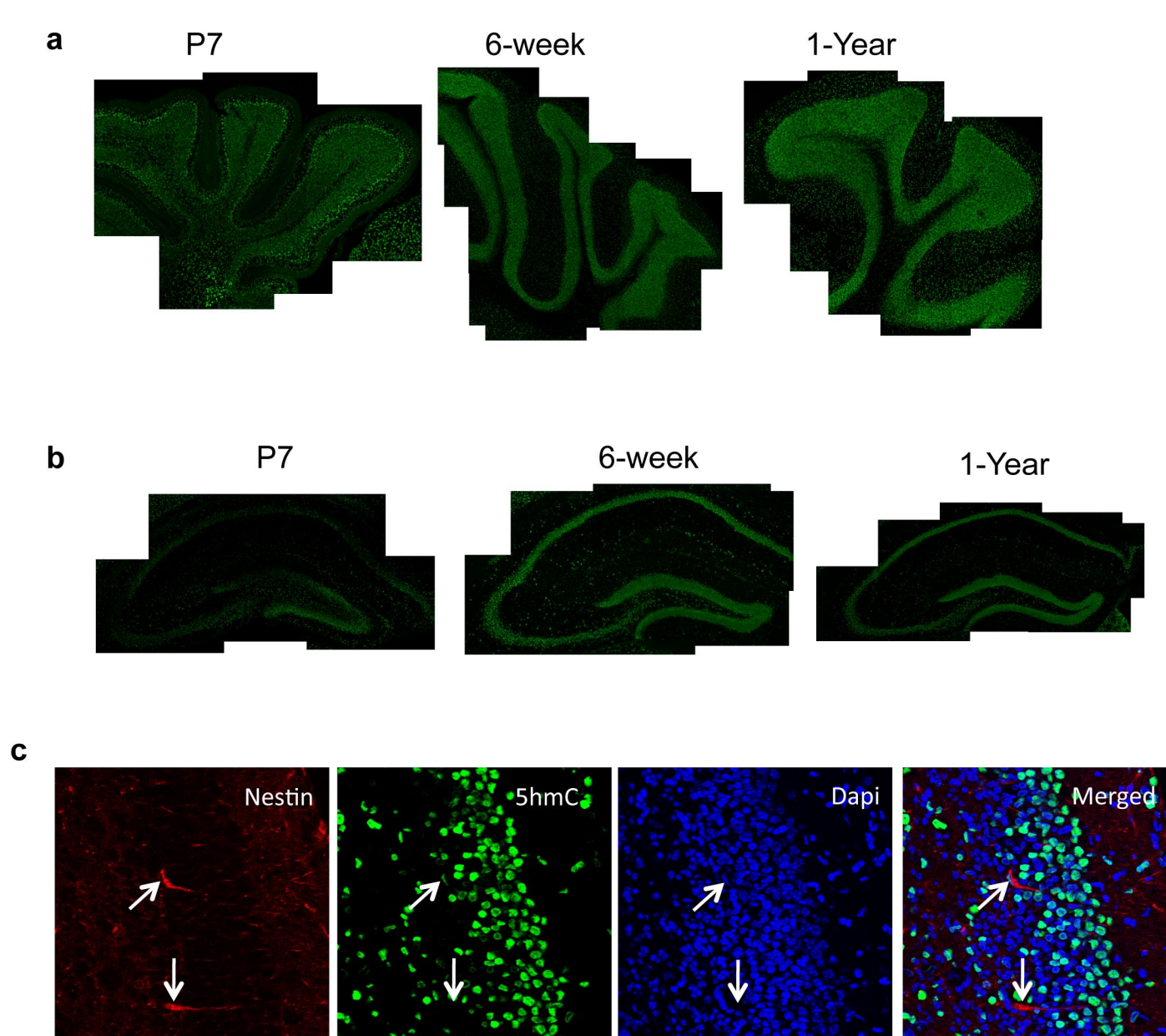
**Supplementary Dataset 6. 6 week Hippocampus DhMRs.** This dataset contains DhMRs identified as enriched for 5-hmC by pairwise comparisons to P7 and 1yr hippocampus as well as P7, 6wk, and 1yr cerebellum.

**Supplementary Dataset 7. 1 year Hippocampus Cerebellum DhMRs.** This dataset contains DhMRs identified as enriched for 5-hmC by pairwise comparisons to P7 and 6wk hippocampus as well as P7, 6wk, and 1yr cerebellum.

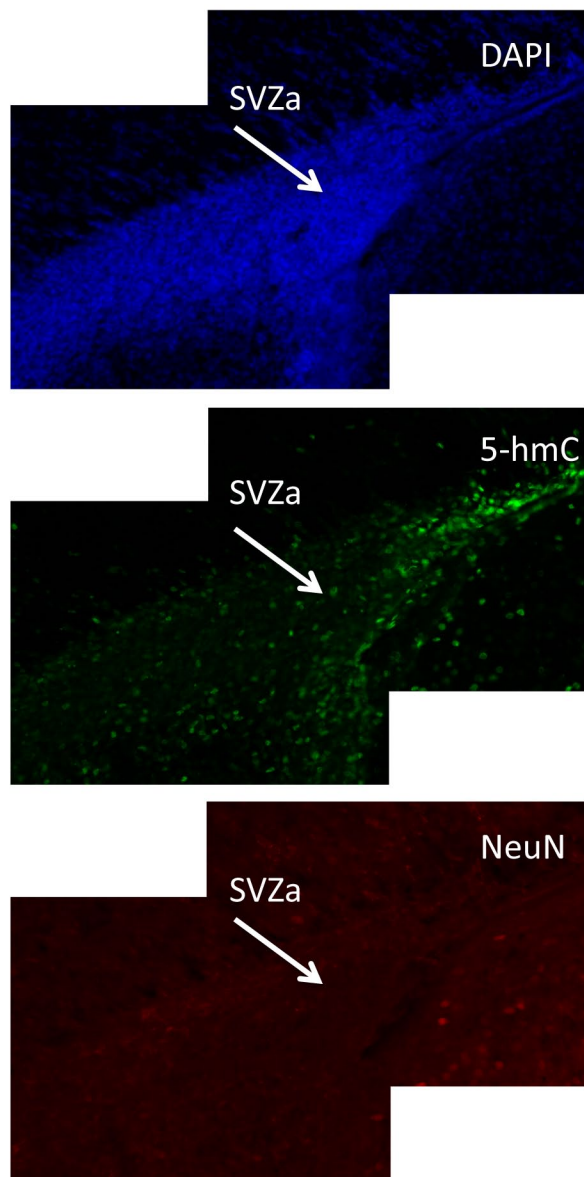
**Supplementary Dataset 8. Tissue specific, Stable, and Dynamic.** This dataset contains cerebellum and hippocampus specific DhMRs as well as the Stable and Dynamic DhMRs summarized in Fig. 3.

**Supplementary Dataset 9. Human 5-hmC enriched regions.** This dataset contains regions identified as enriched for 5-hmC in both human samples tested.

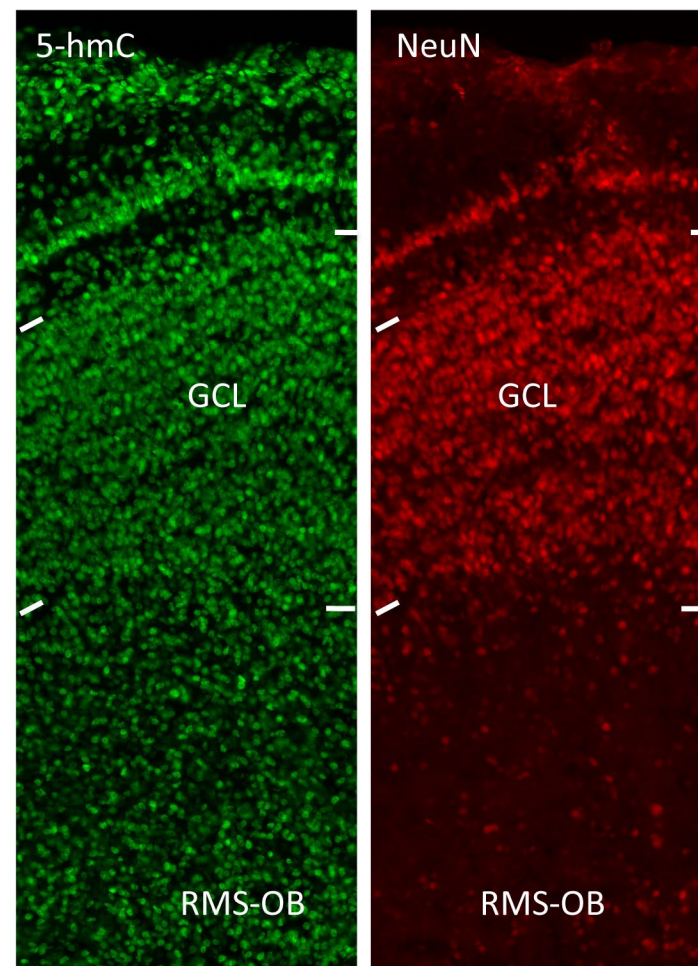
**Supplementary Dataset 10. Gene Ontology Analyses.** This dataset contains GO analyses for 1.) genes exhibiting activated gene expression in 6wk Cb relative to P7 Cb also with 5-hmC enriched in 6wk Cb relative to P7 Cb (“Adult\_activated\_5hmC\_GO”) 2.) genes with overlapping Adult dynamic DhMRs in both cerebellum and hippocampus (“Adult\_dynamic\_DhMR\_GO”) 3.) genes with overlapping Adult stable DhMRs in both cerebellum and hippocampus (“Adult\_stable\_DhMR\_GO”) and 4.) genes with overlapping 1yr specific DhMRs in both cerebellum and hippocampus (“1yr\_specific\_DhMR\_GO”).



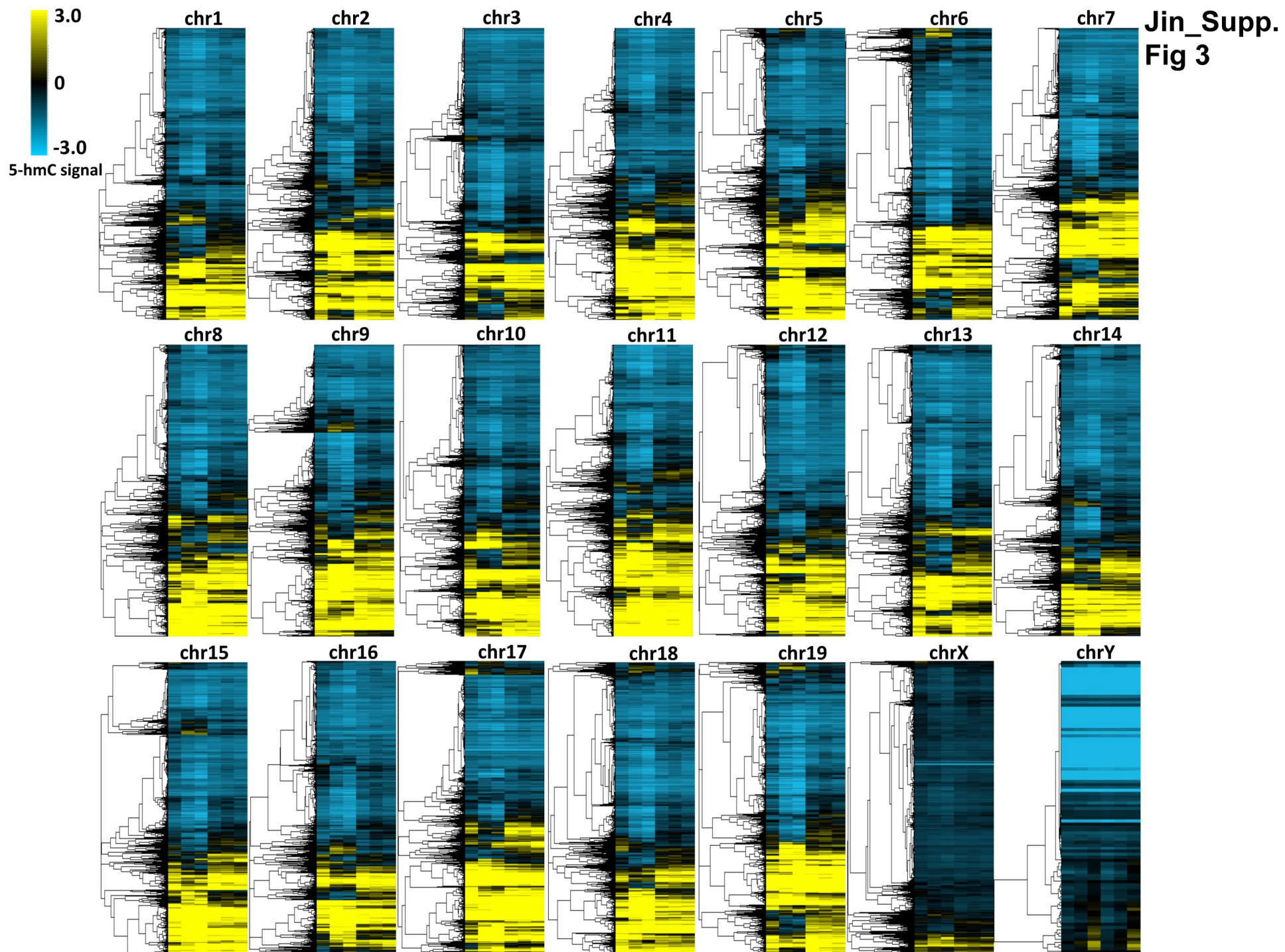
**a**

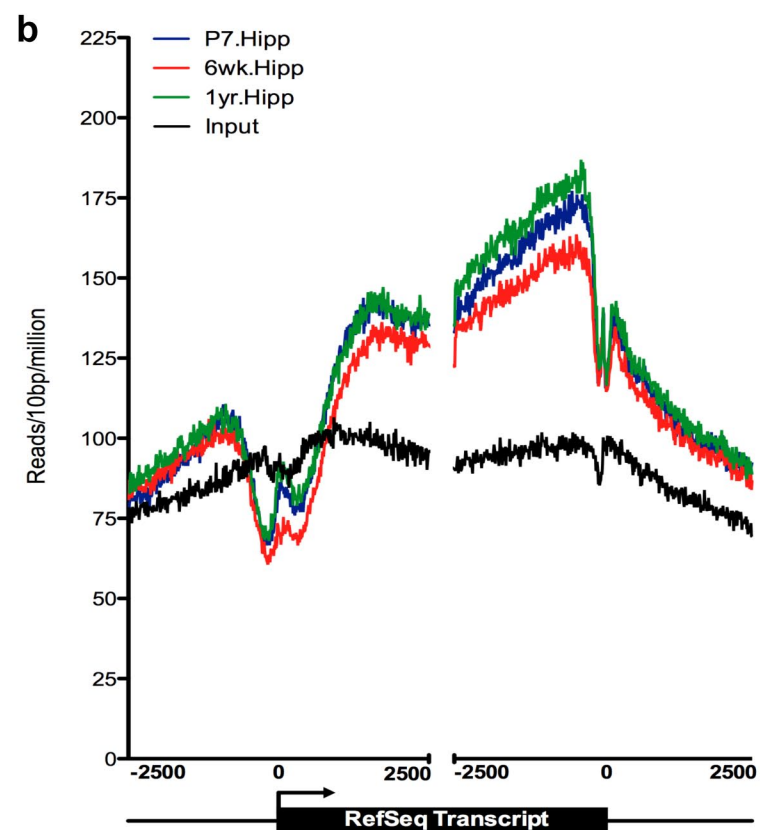
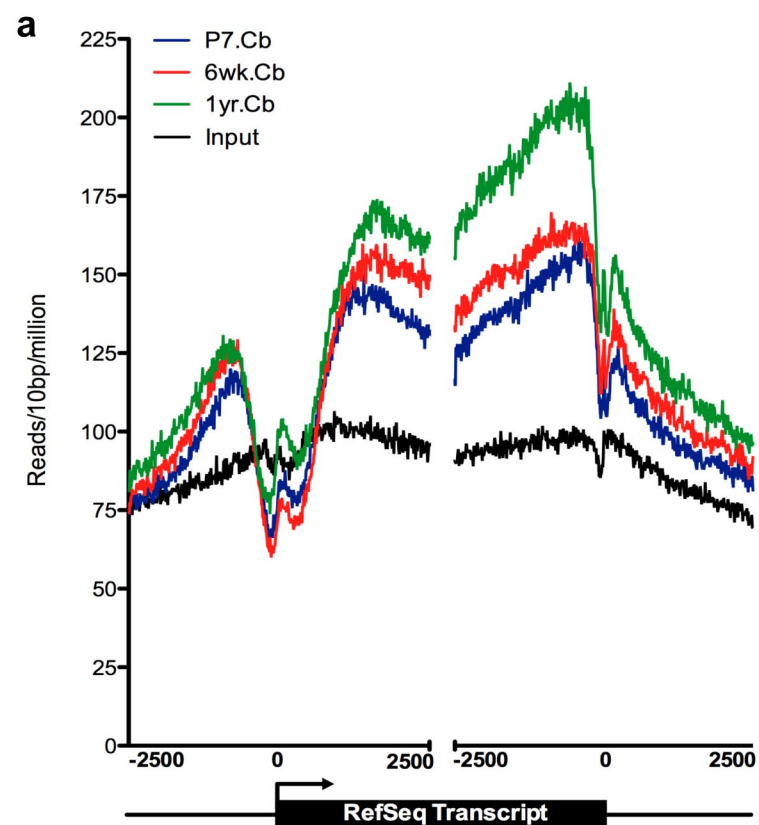


**b**

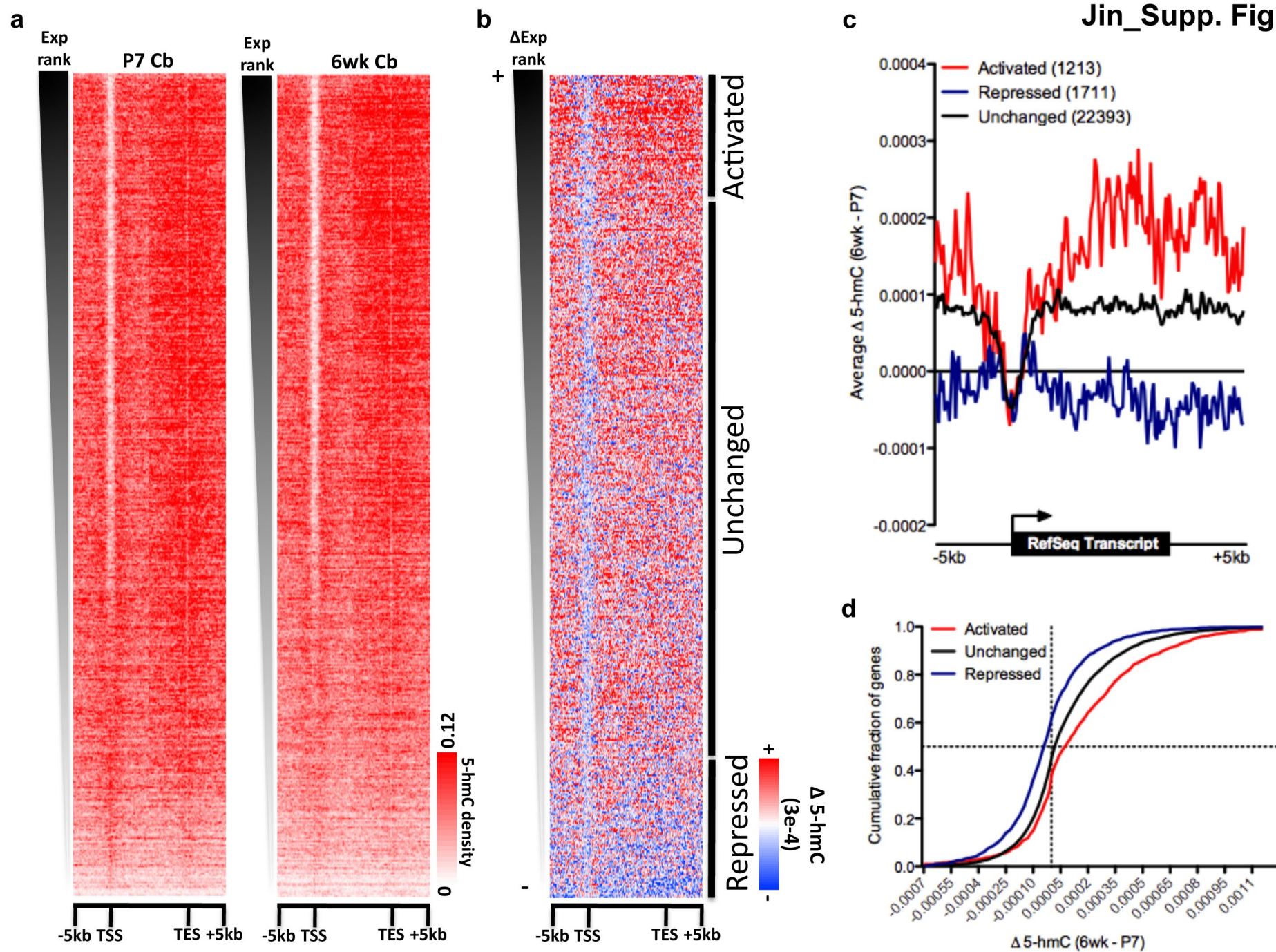


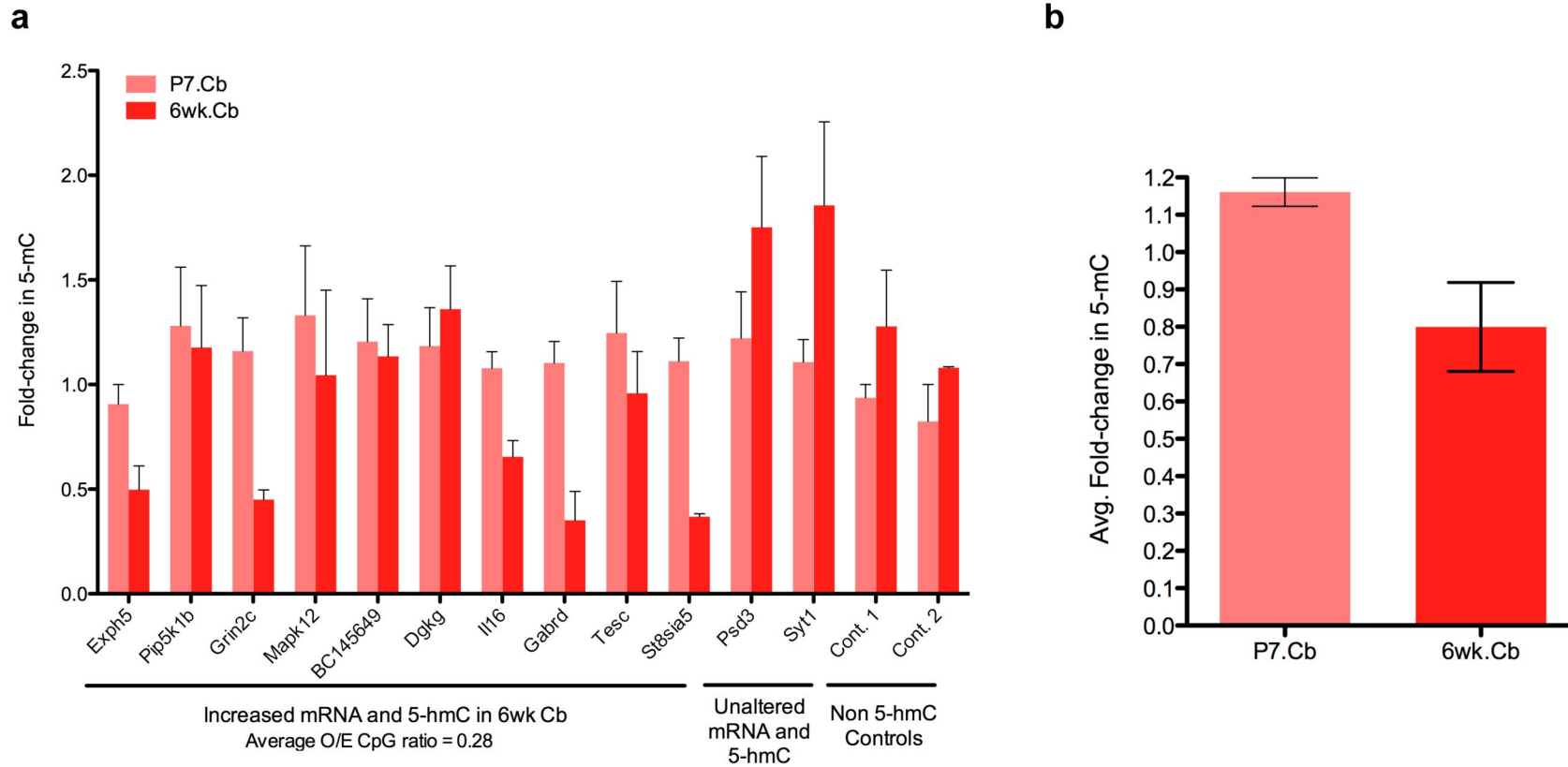


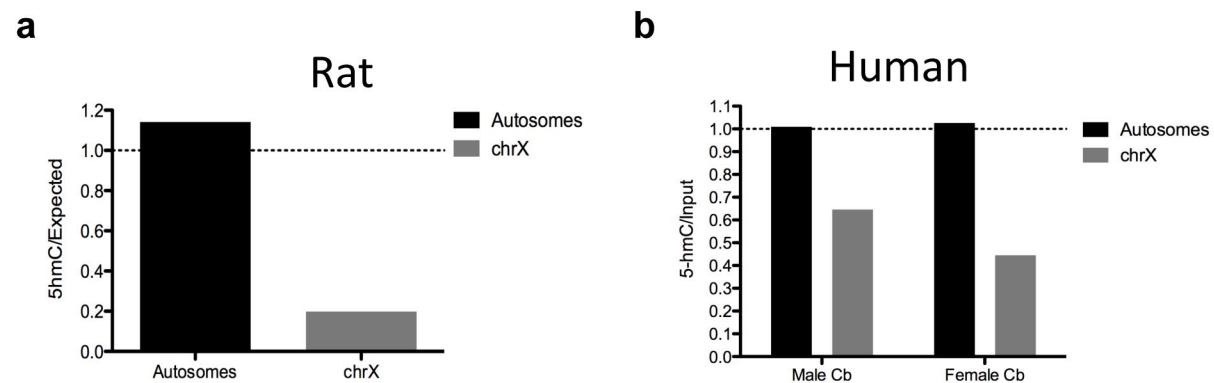




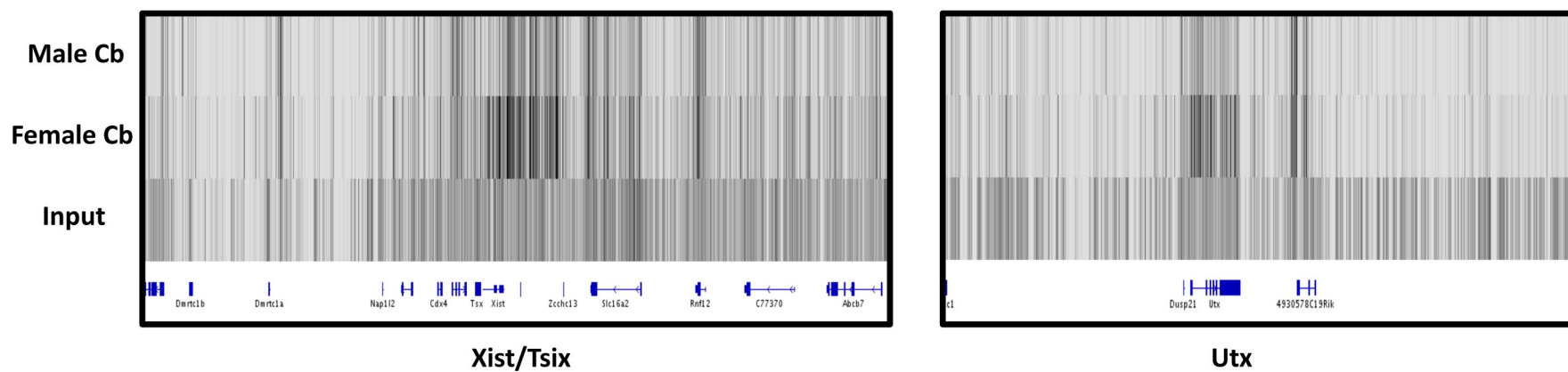


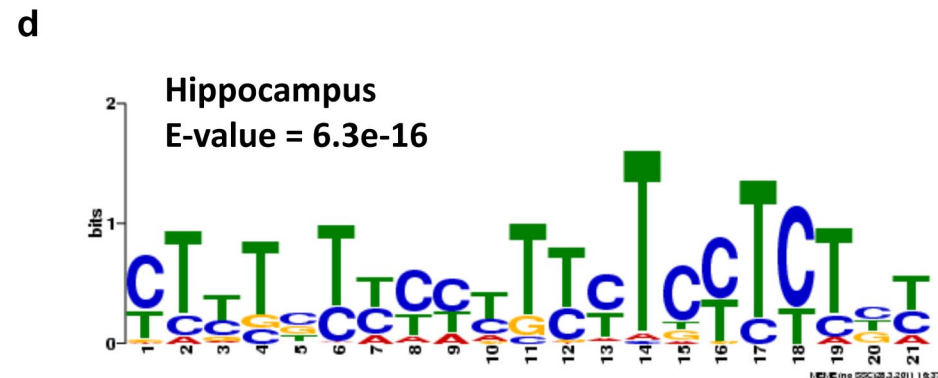
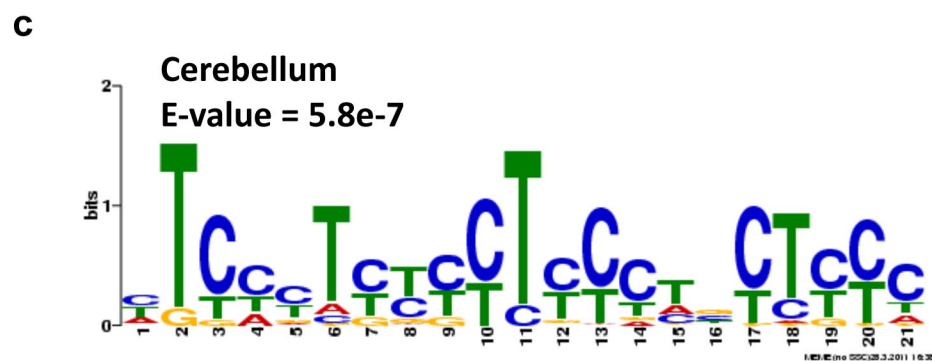
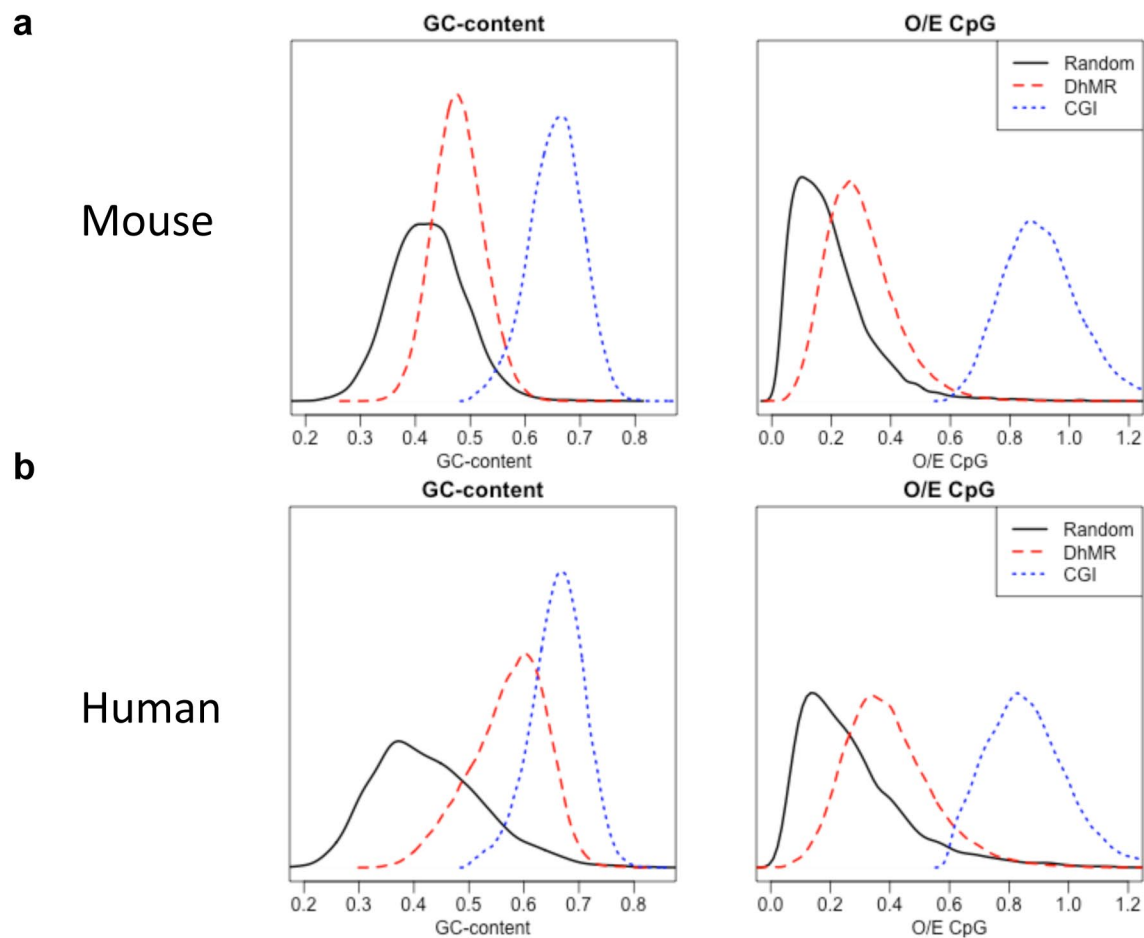




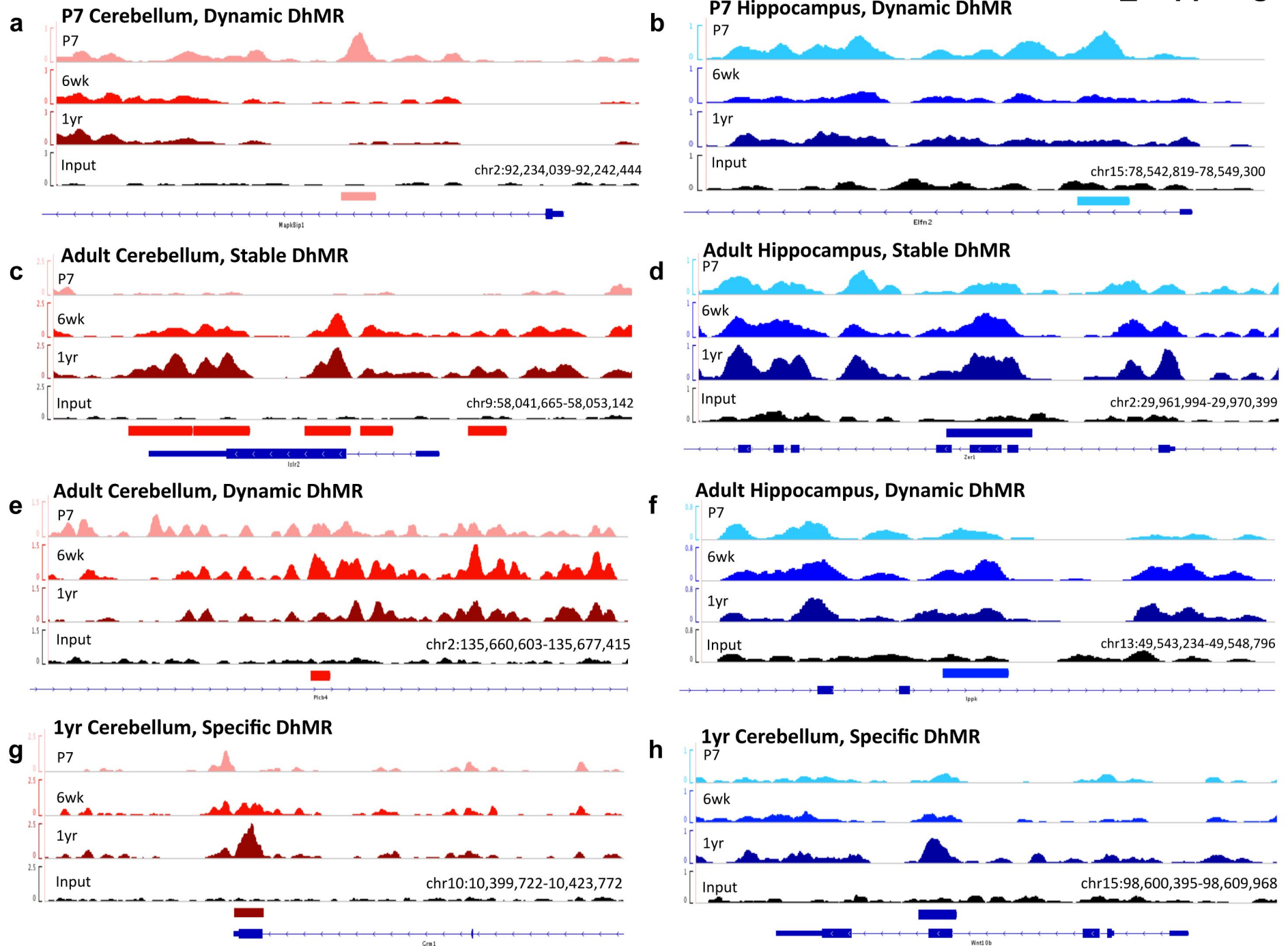


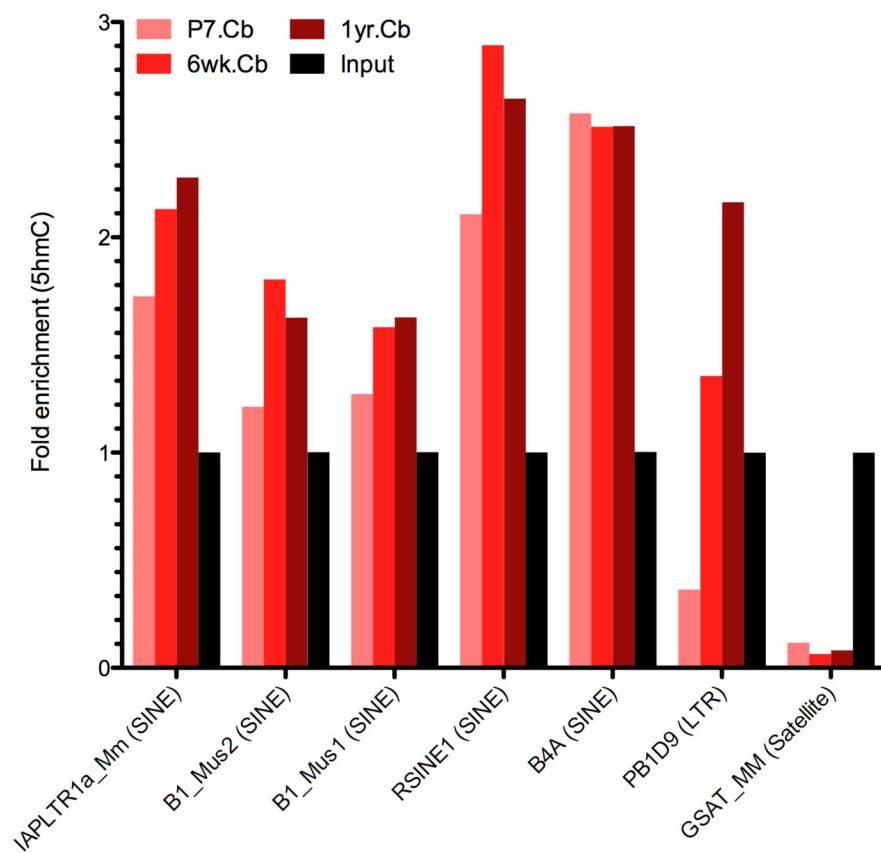
**c**









**a****b**